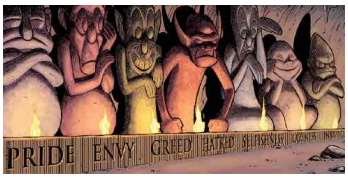
data blueprint
UNLOCKING BUSINESS VALUE



Exorcising
the Seven Deadly Data Sins

Peter Aiken, PhD

datablueprint.com

Peter Aiken, Ph.D.

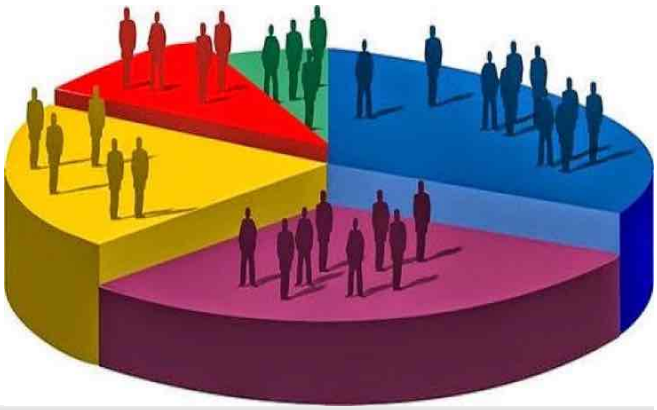
- 33+ years in data management
- Repeated international recognition
- Founder, Data Blueprint (datablueprint.com)
- Associate Professor of IS (vcu.edu)
- DAMA International (dama.org)
- 10 books and dozens of articles
- Experienced w/ 500+ data management practices
- Multi-year immersions:
  - US DoD (DISA/Army/Marines/DLA)
  - Nokia
  - Deutsche Bank
  - Wells Fargo
  - Walmart
  - …

Disclaimer

- This presentation at this event does not indicate endorsement of any specific method, process, or certification!

Census

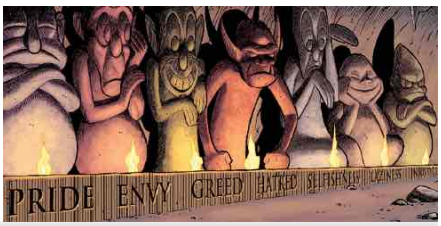
- How many self identify as primarily business?
- How many self identify as IT?
- How many as both?
- How many are satisfied with their ability to leverage data in support of the organizational mission?



Exorcising the Seven Deadly Data Sins

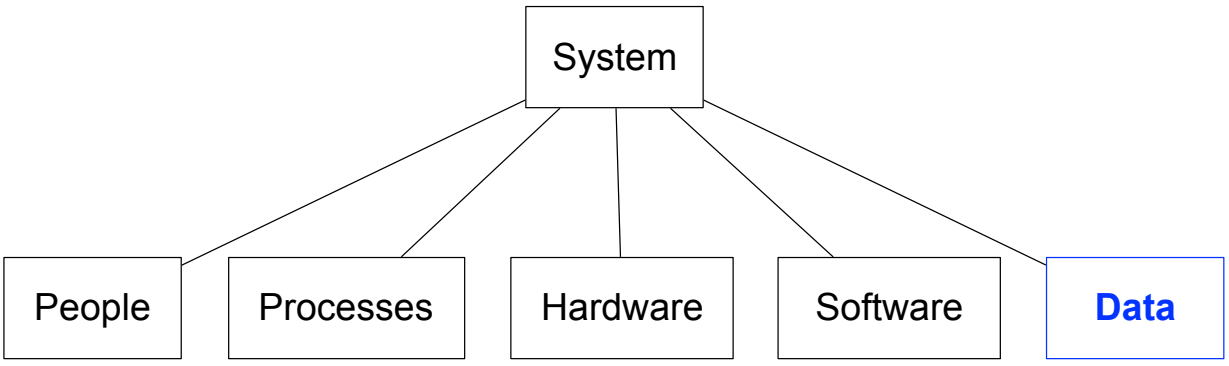
Context

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges



What is a system?

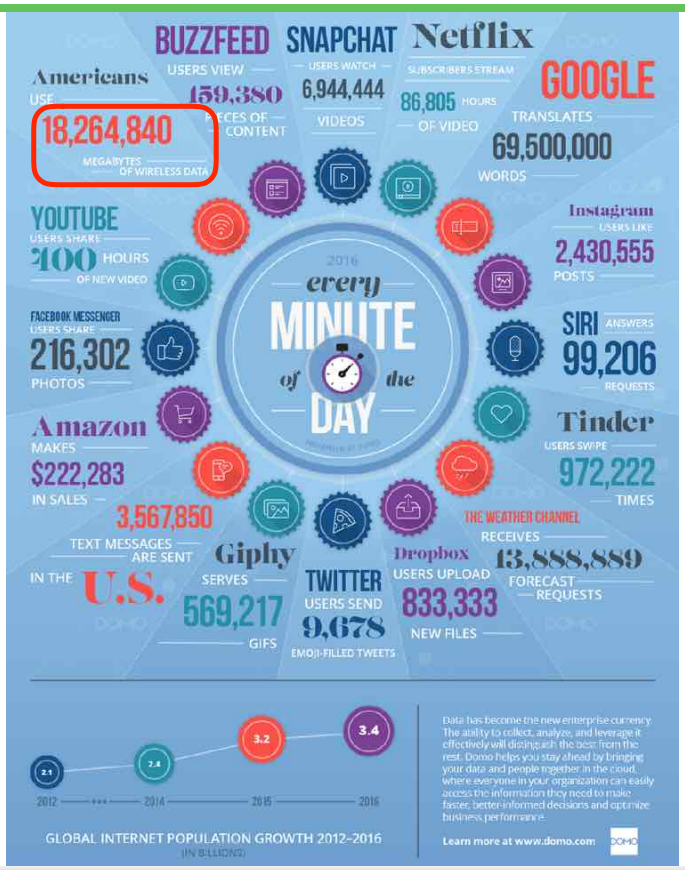
- A set of detailed methods, procedures, and routines established or formulated to carry out a specific activity, perform a duty, or solve a problem.
- An organized, purposeful structure regarded as a whole and consisting of interrelated and interdependent elements (components, entities, factors, members, parts etc.). These elements continually influence one another (directly or indirectly) to maintain their activity and the existence of the system, in order to achieve the goal of the system.



http://www.businessdictionary.com/definition/system.html#ixzz23T7LyAjJ

As articulated by Micheline Casey

How much data was generated every minute in 2016!



Organizational Assets

- Cash & other financial instruments
- Real property
- Inventory
- Intellectual Property
- Human
  - **K**nowledge
  - **S**kills
  - **A**bilities
- Financial
- Organizational reputation
- Good will
- Brand name
- **Data**!!!

Data Assets Win!

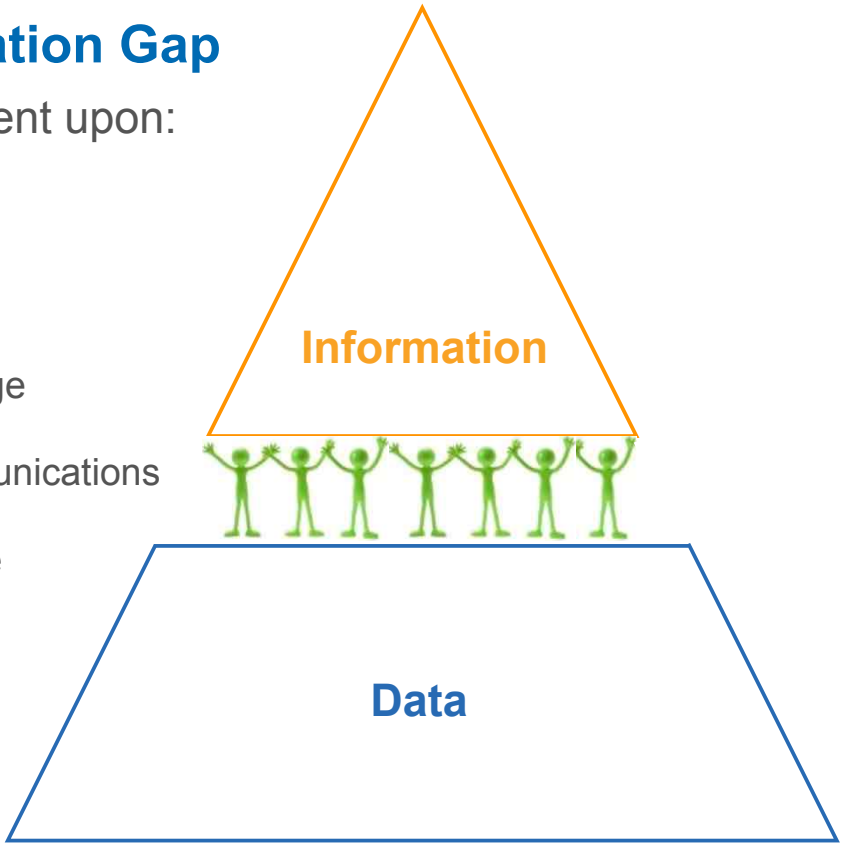
Asset: A resource controlled by the organization as a result of past events or transactions and from which future economic benefits are expected to flow [Wikipedia]

- Today, data is the most powerful, yet underutilized and poorly managed organizational asset
- Data is your
  - Sole
  - Non-depletable
  - Non-degrading
  - Durable
  - Strategic
- Asset
  - Data is the new oil!
  - Data is the new (s)oil!
  - Data is the new bacon!
- As such, data deserves:
  - It's own strategy
  - Attention on par with similar organizational assets
  - Professional ministration to make up for past neglect

	Data Assets	Financial Assets	Real Estate Assets	Inventory Assets
Non-depletable	Available for subsequent use	Can be used up		Can be used up
Non-degrading	√	√	Can degrade over time	Can degrade over time
Durable	Non-taxed		√	√
Strategic Asset	√	√	√	√

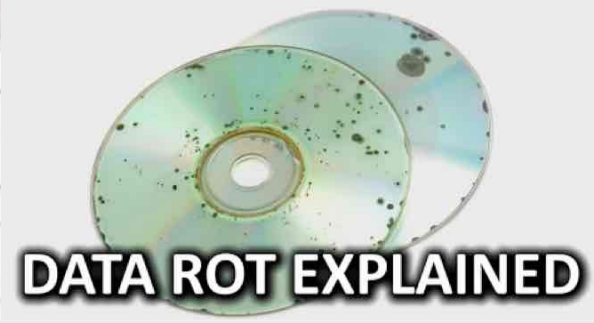
Data / Information Gap

- Overly dependent upon:
  - Human-beings
  - Wetwear
  - Tribal knowledge
  - Informal communications
  - Non-repeatable practices



Separating the Wheat from the Chaff

- Data that is better organized increases in value
- Poor data management practices are costing organizations much money/time/effort
- 80% of organizational data is ROT
  - **R**edundant
  - **O**bsolete
  - **T**rivial



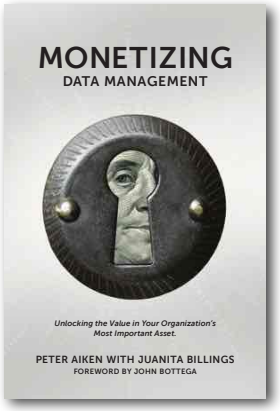
Put simply, organizations:

- Have little idea what data they have
- Do not know where it is (and)

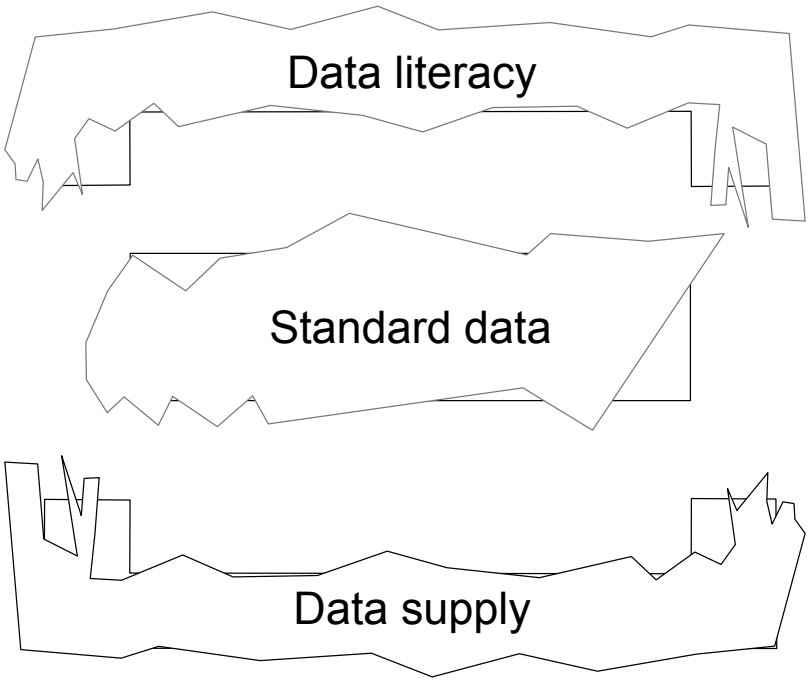
- Do not know what their knowledge workers do with it

Data is a hidden IT Expense

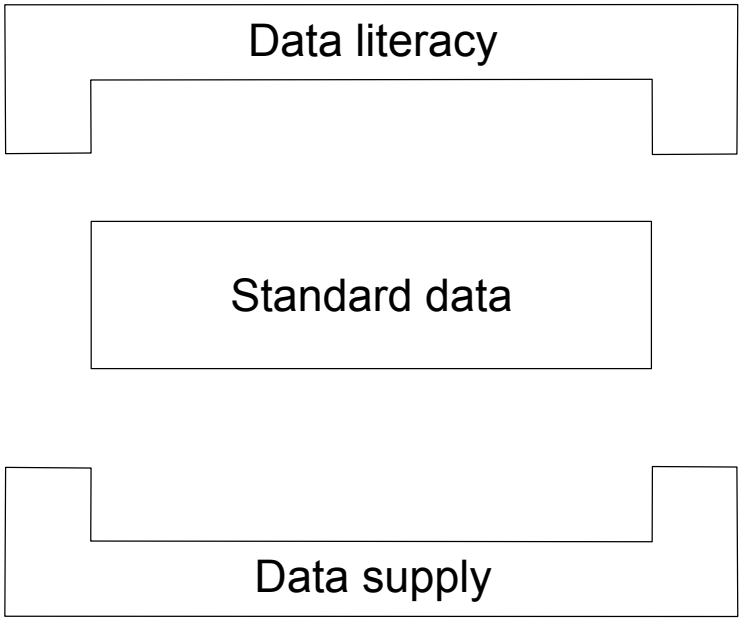
- Organizations spend between 20 - 40% of their IT budget evolving their data - including:
  - Data **migration**
    - Changing the location from one place to another
  - Data **conversion**
    - Changing data into another form, state, or product
  - Data **improving**
    - Inspecting and manipulating, or re-keying data to prepare it for subsequent use
      - Source: *John Zachman*



Making a Better Data Sandwich

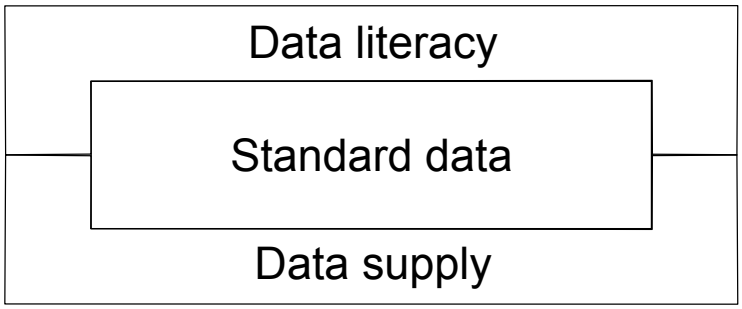


Making a Better Data Sandwich



Making a Better Data Sandwich

This cannot happen without engineering and architecture!



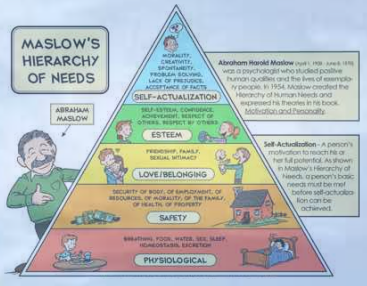
Quality engineering/architecture work products do not happen accidentally!

Data Management Practices Hierarchy

You can accomplish Advanced Data Practices without becoming proficient in the Foundational Data Practices however this will:

- Take longer
- Cost more
- Deliver less
- Present greater risk

(with thanks to Tom DeMarco)



Technologies

Advanced Data Practices

- MDM
- Mining
- Big Data
- Analytics
- Warehousing
- SOA

Foundational Data Practices

Data Governance

Data Quality

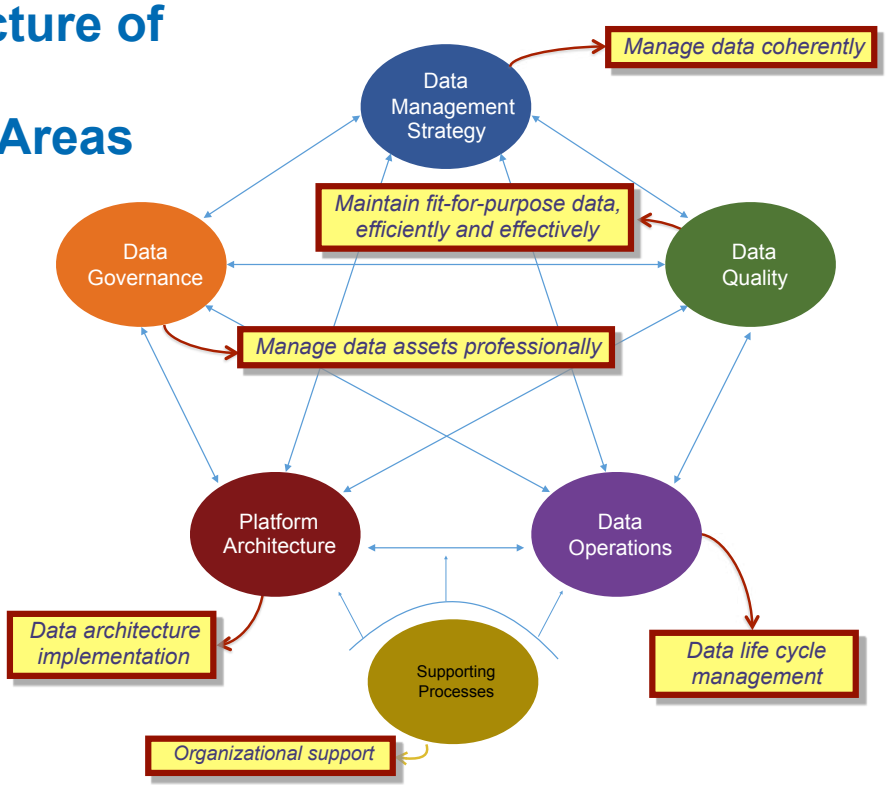
Data Management Strategy

Data Platform/Architecture

Data Operations

Capabilities

DMM℠ Structure of 5 Integrated DM Practice Areas



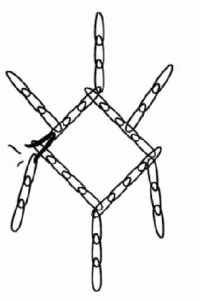
CMMI Institute

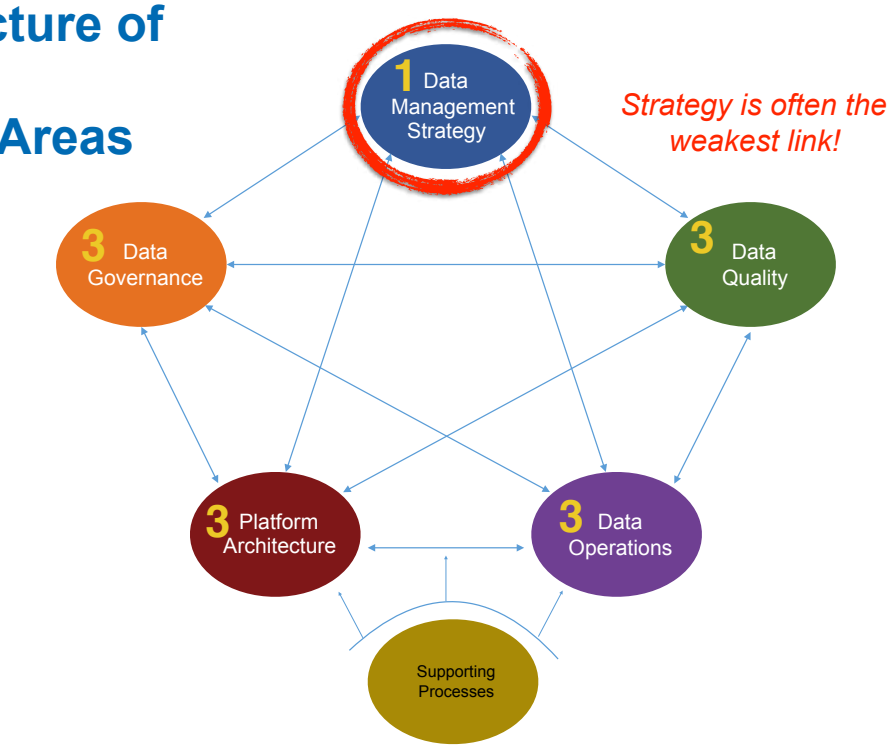
data blueprint

UNLOCKING BUSINESS VALUE

DMM℠ Structure of 5 Integrated DM Practice Areas

5
4
3
2
1



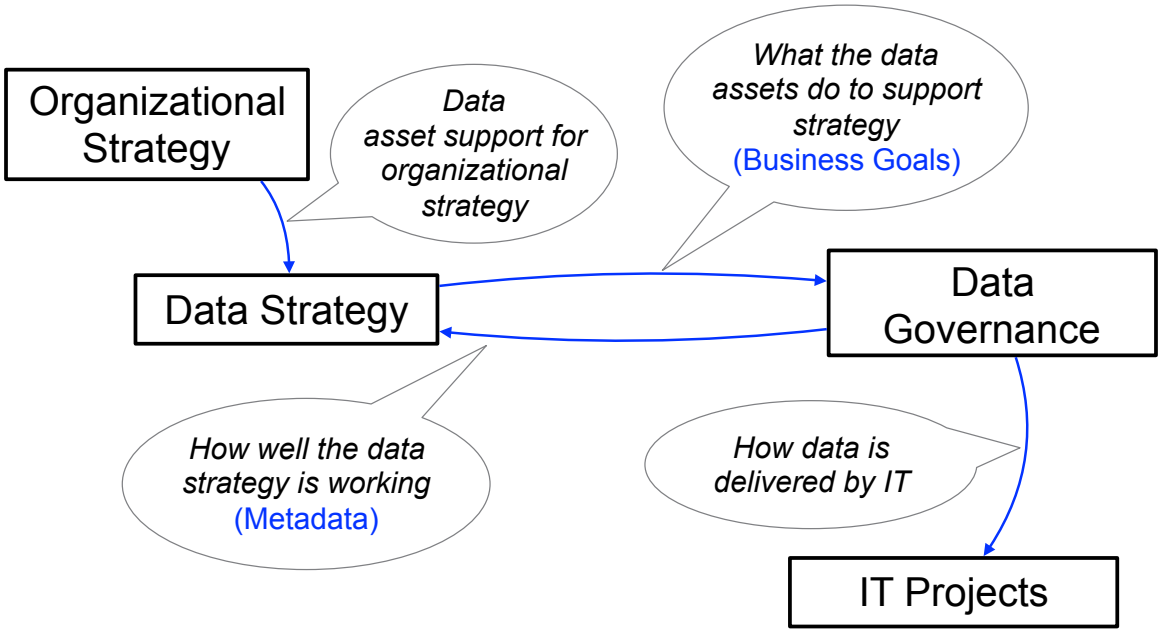


CMMI Institute

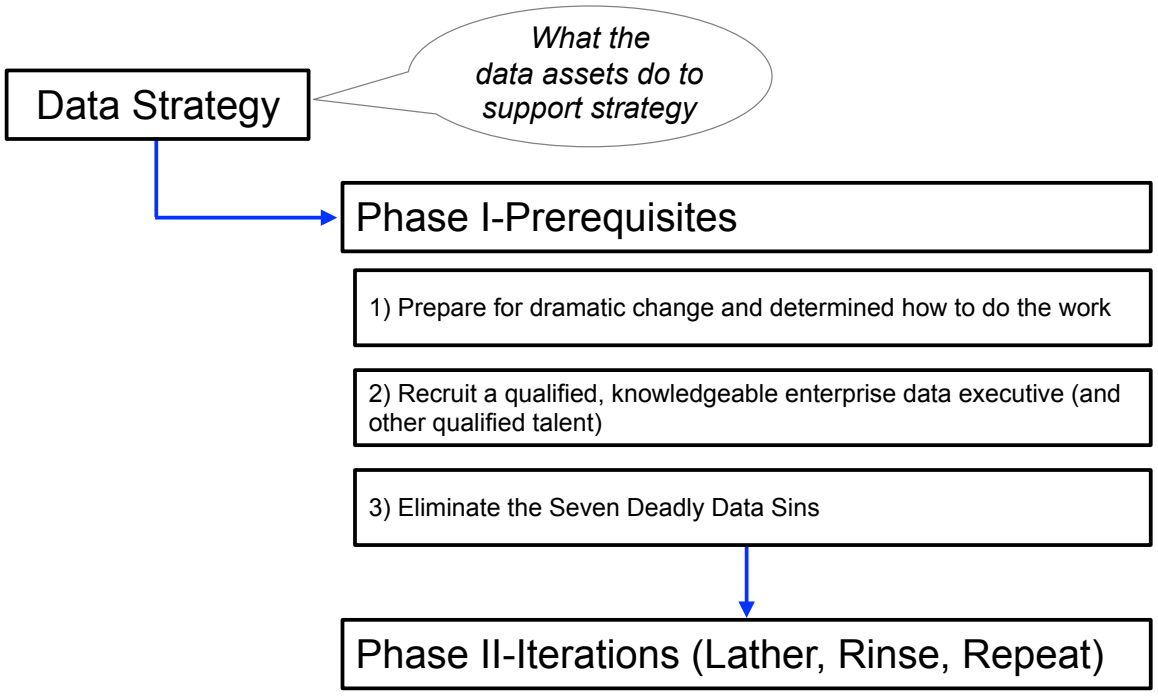
data blueprint

UNLOCKING BUSINESS VALUE

Data Strategy and Data Governance in Context



Data Strategy is Implemented in 2 Phases



Data Strategy is Implemented in 2 Phases

What the data assets do to support strategy

Data Strategy

Phase I-Prerequisites

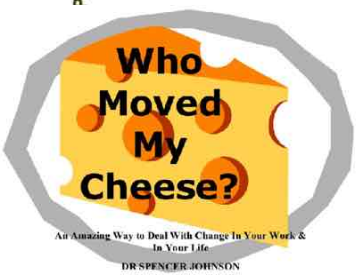
1) Prepare for dramatic change and determine how to do the work

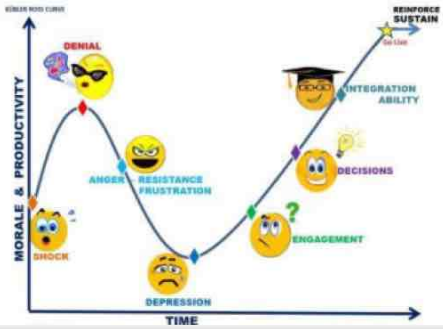
2) Recruit a qualified, knowledgeable enterprise data executive (and other qualified talent)

3) Eliminate the Seven Deadly Data Sins

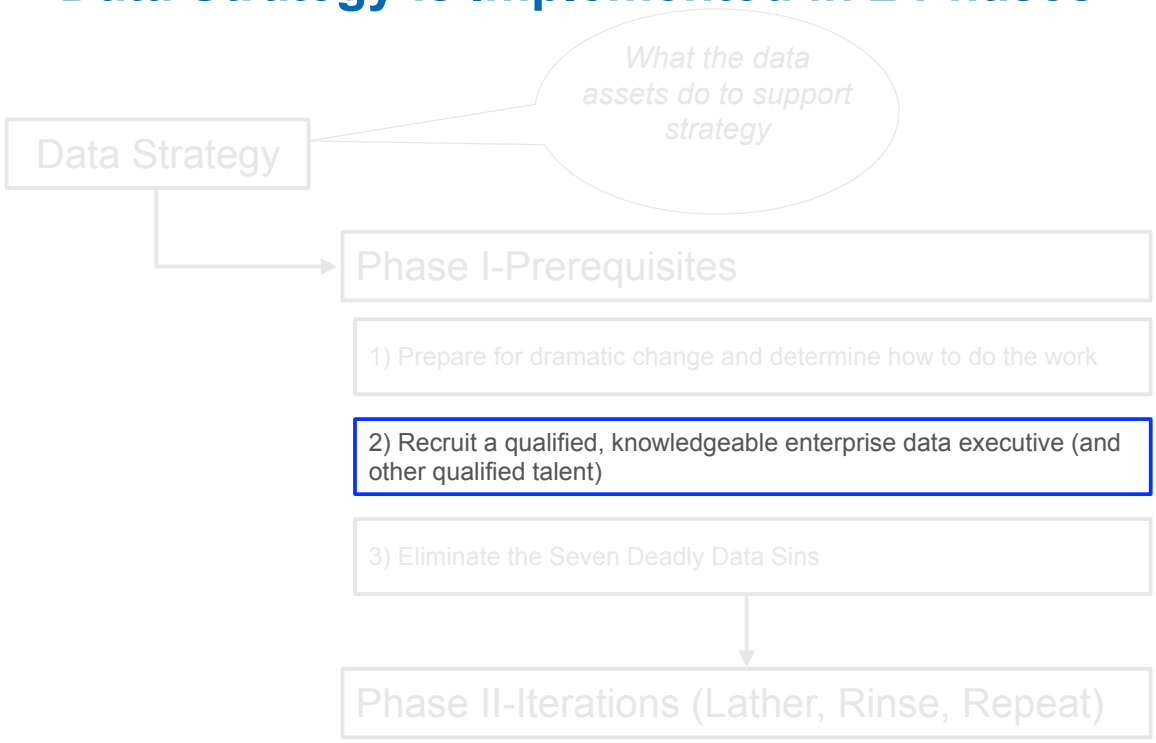
Phase II-Iterations (Lather, Rinse, Repeat)

Change Management

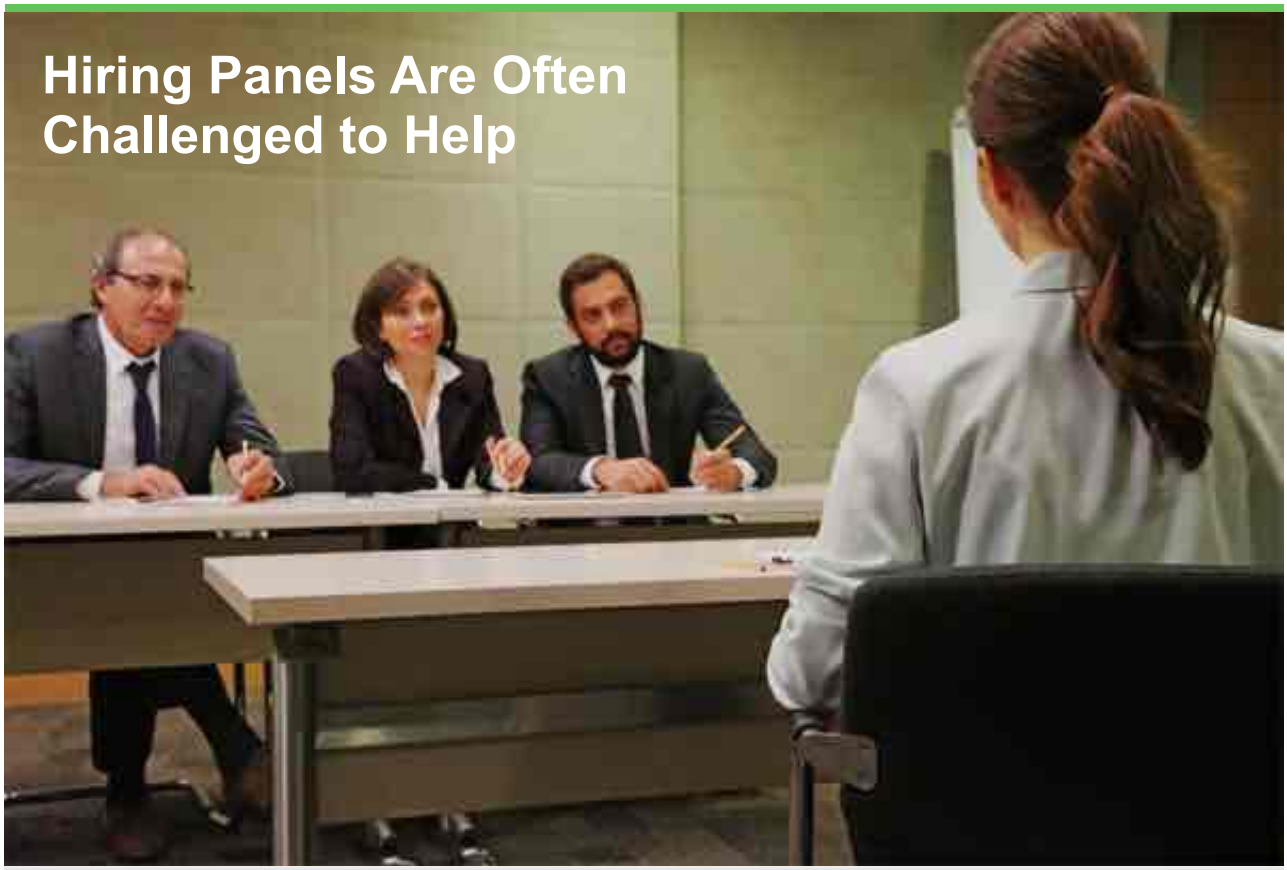




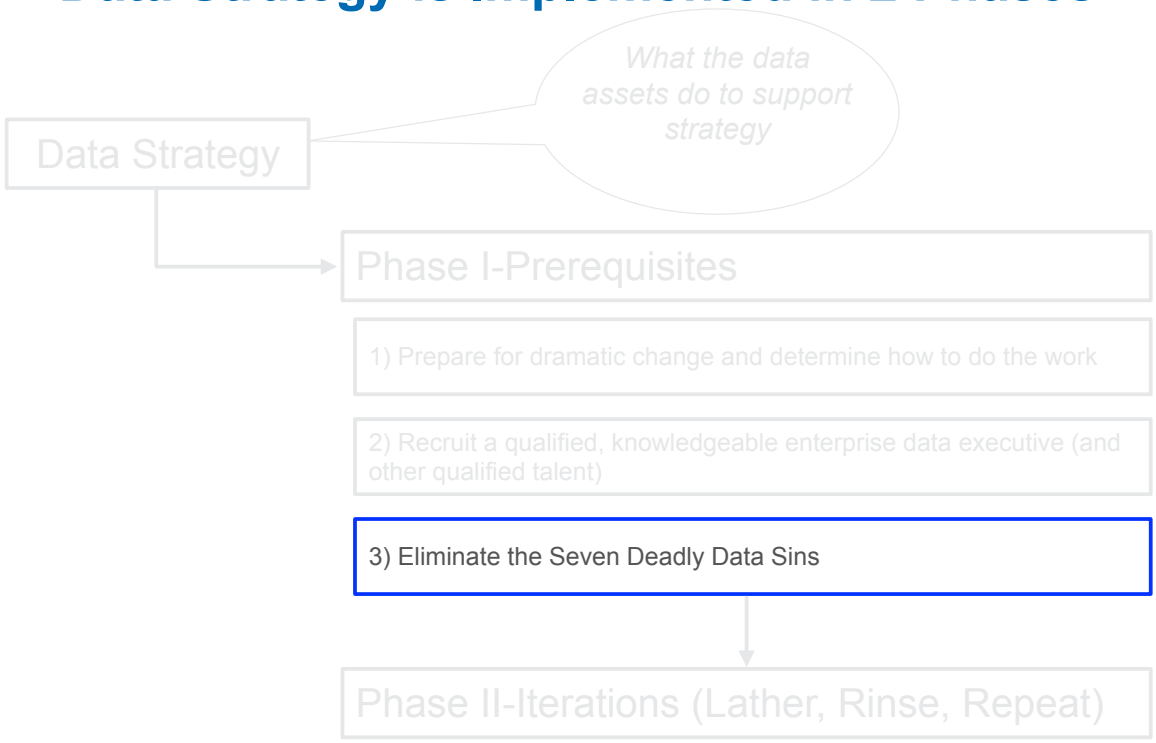
Data Strategy is Implemented in 2 Phases



Hiring Panels Are Often Challenged to Help



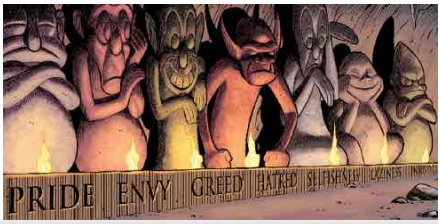
Data Strategy is Implemented in 2 Phases



Exorcising the Seven Deadly Data Sins

Context

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges



George Box
British Statistician
(1919-2013)

"All models are wrong, ...
... some are useful."

theDataDoctrine.com

DATA centric

We are uncovering better ways of developing IT systems by doing it and helping others do it. Through this work we have come to value:

Data programmes preceding software development
Stable data structures preceding stable code
Shared data preceding completed software
Data reuse preceding reusable code

That is, while there is value in the items on the right, we value the items on the left more.

Data programmes preceding software development

Data programmes preceding software development

Common Organizational Data
(and corresponding data needs requirements)

Future State

(Version +1)

Evolve

Data evolution is separate from, external to, and precedes system development life cycle activities!

Data management and software development must be separated and sequenced

Systems Development Activities

Create

New Organizational Capabilities

theDataDoctrine.com

We are uncovering better ways of developing IT systems by doing it and helping others do it. Through this work we have come to value:

Data programmes preceding software development

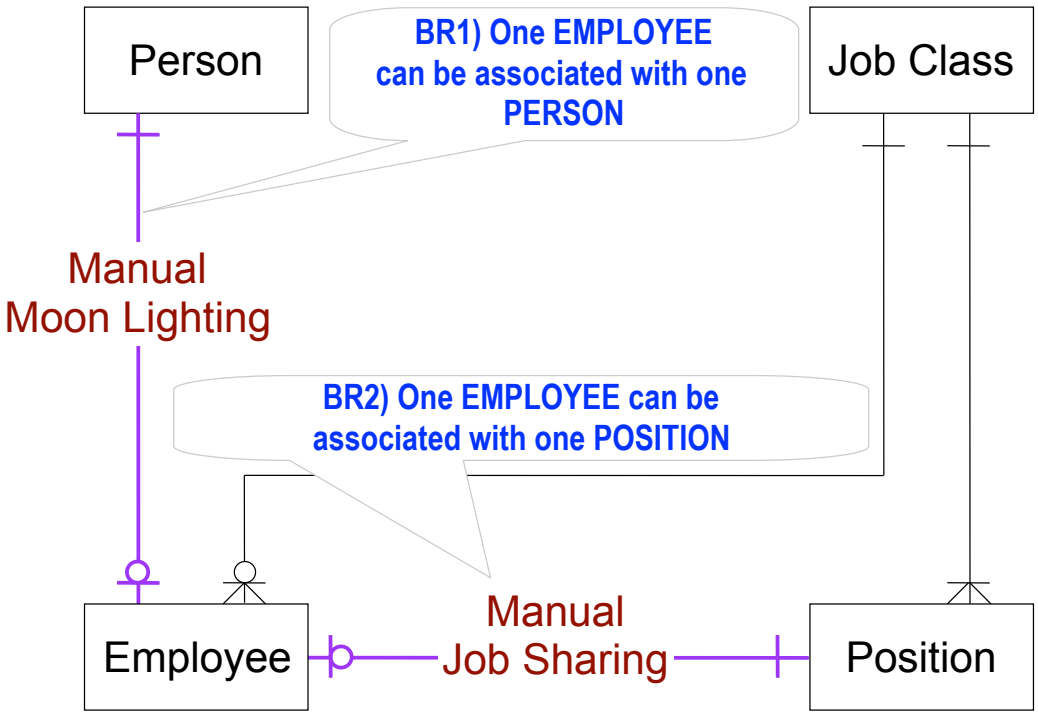
Stable data structures preceding stable code

Shared data preceding completed software

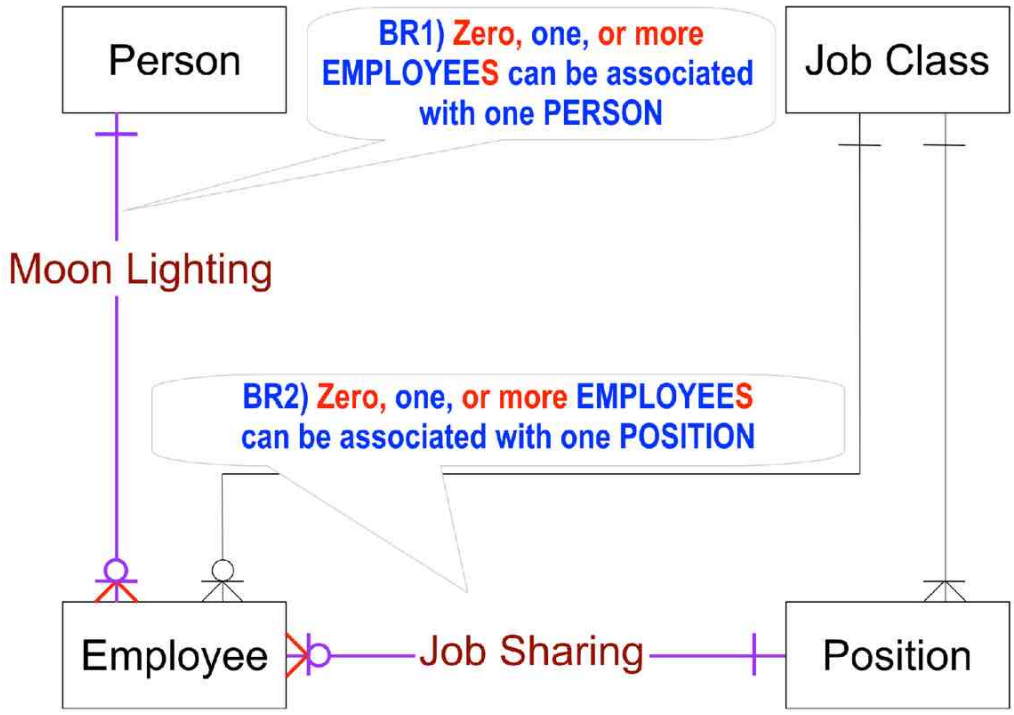
Data reuse preceding reusable code

That is, while there is value in the items on the right, we value the items on the left more.

Stable data structures preceding stable code

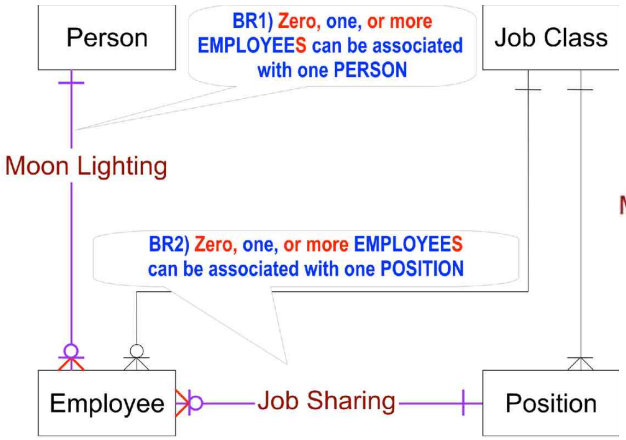


Stable data structures preceding stable code

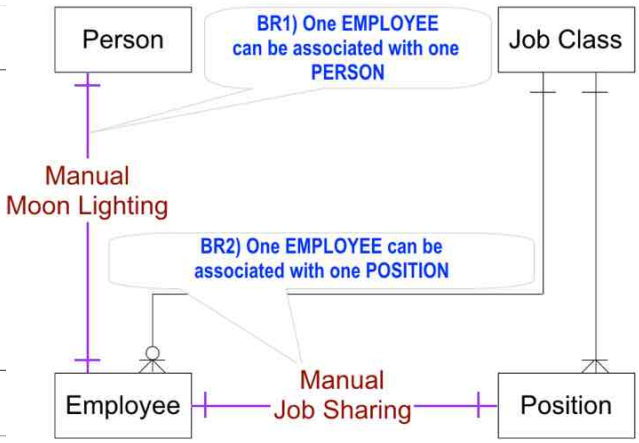


Stable data structures preceding stable code

More flexible data structure



Less flexible data structure



(Requires 2 structural loops more than the more flexible data structure)

Data structures must be specified prior software development/acquisition

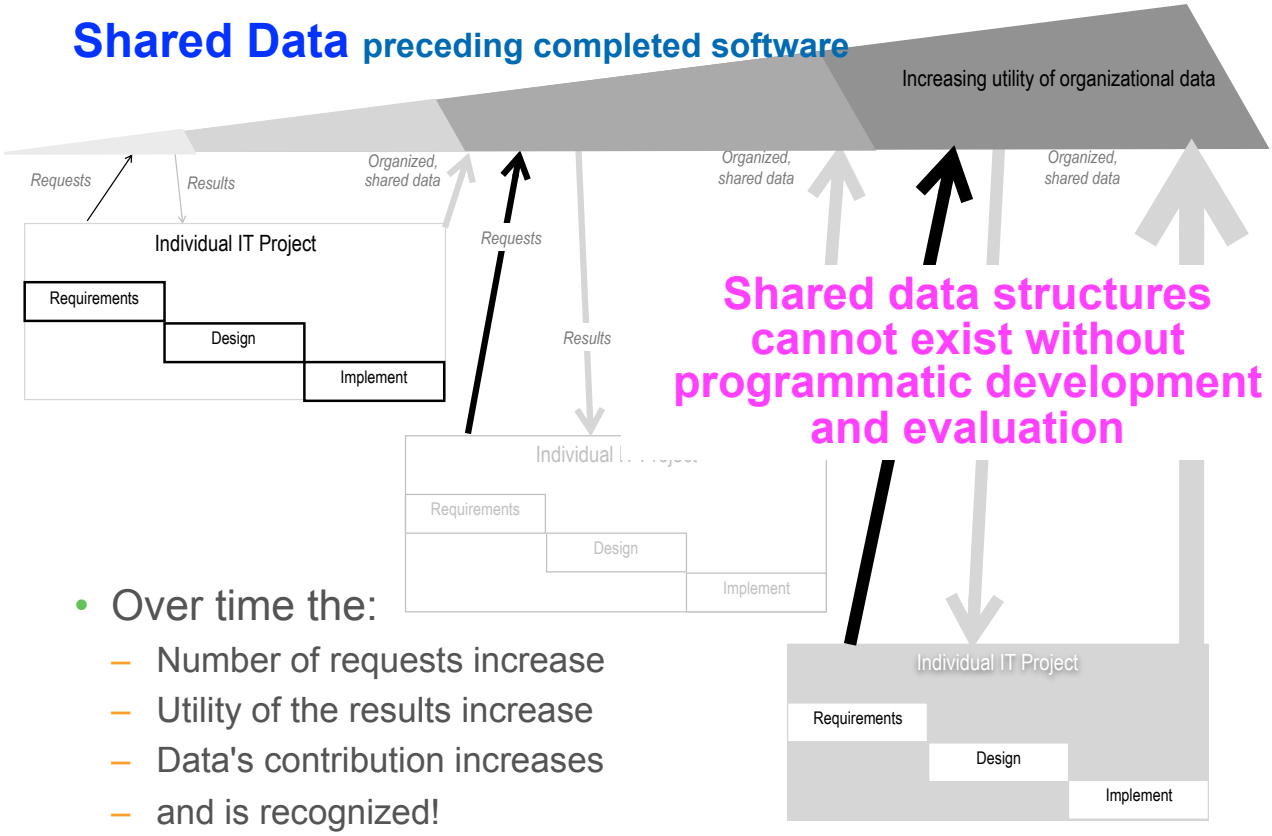
theDataDoctrine.com

We are uncovering better ways of developing IT systems by doing it and helping others do it. Through this work we have come to value:

Data programmes preceding software development
Stable data structures preceding stable code
Shared data preceding completed software
Data reuse preceding reusable code

That is, while there is value in the items on the right, we value the items on the left more.

Shared Data preceding completed software



- Over time the:
  - Number of requests increase
  - Utility of the results increase
  - Data's contribution increases
  - and is recognized!

theDataDoctrine.com

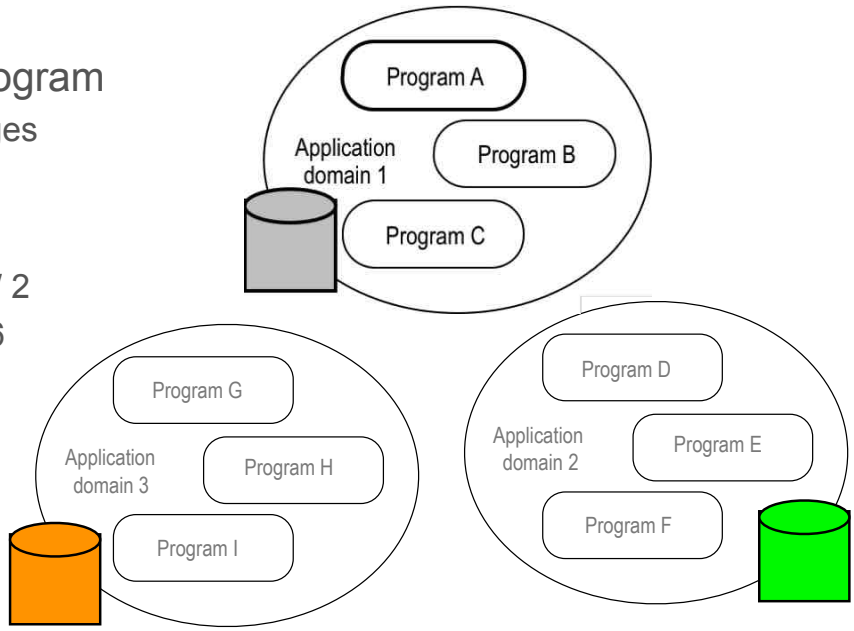
We are uncovering better ways of developing IT systems by doing it and helping others do it. Through this work we have come to value:

Data programmes preceding software development
Stable data structures preceding stable code
Shared data preceding completed software
Data reuse preceding reusable code

That is, while there is value in the items on the right, we value the items on the left more.

Data reuse preceding reusable code

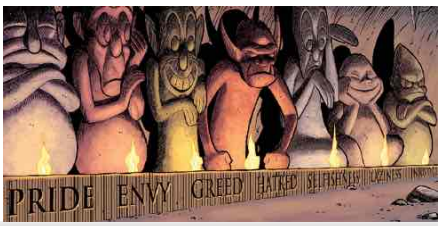
- Reusable software has been valued more than reusable data
- Who makes decisions about the range and scope of common data usage?
- Change a program
  - 9 max changes
- Change data
  - Worst case
  - (N * (N - 1)) / 2
  - (9 * 8)/2 = 36



Exorcising the Seven Deadly Data Sins

Context

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges



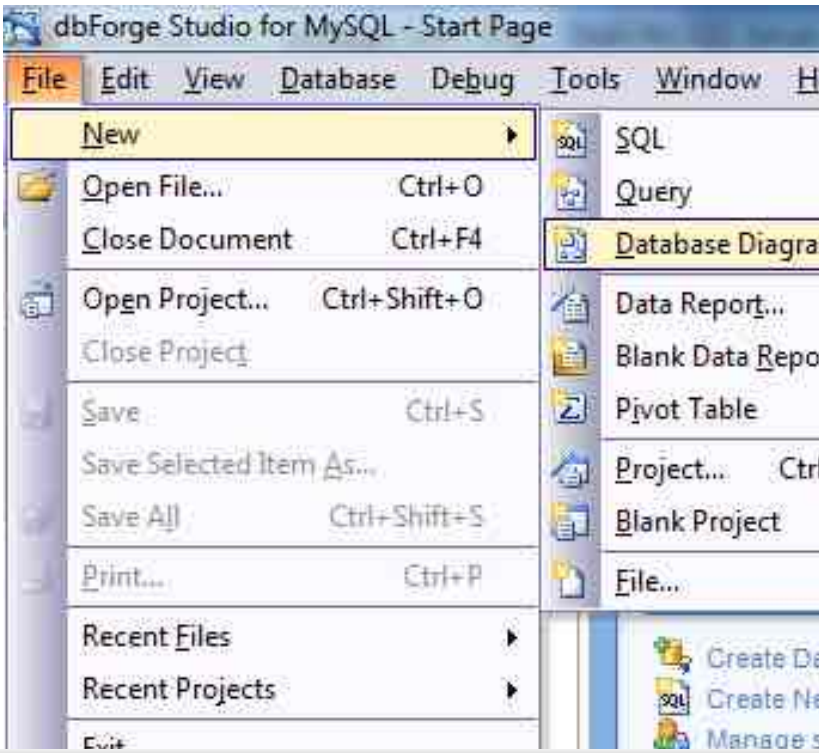
What do we teach knowledge workers about data?

What percentage of the deal with it daily?

100%

What do we teach IT professionals about data?

- 1 course
  - How to build a new database
- What impressions do IT professionals get from this education?
  - Data is a technical skill that is needed when developing new databases



If the only tool you know is a hammer you tend to see every problem as a nail (slightly reworded from Abraham Maslow)

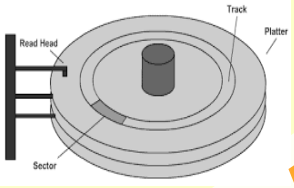
Example

Calculate the access time for a disk with 512 byte/sector and 12 ms advertised seek time. The disk rotates at 5400 RPM and transfers data at a rate of 4MB/sec. The controller overhead is 1 ms. Assume that the queue is idle (so no service time)

Answer:

Disk Access Time = Seek time + Rotational Latency + Transfer time + Controller Time + Queuing Delay

= 12 ms + 0.5 / 5400 RPM + 0.5 KB / 4 MB/s + 1 ms + 0



= 12 ms + 0.5/90 RPS + 0.125 / 1024 s + 1 ms + 0

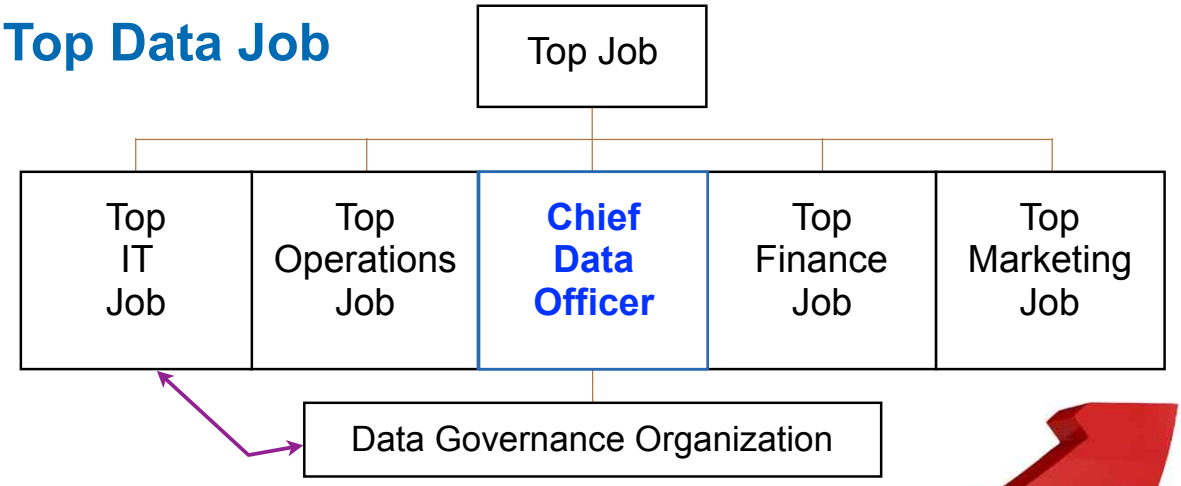
= 12 ms + 5.5 ms + 0.1 ms + 1 ms + 0 ms

= 18.6 ms

If real seeks are 1/3 the advertised seeks, disk access time would be 10.6 ms, with rotation delay contributing 50% of the access time!

The disks no longer rotate!

Top Data Job

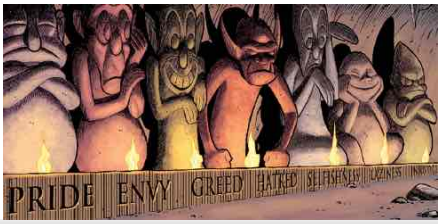


- Dedicated solely to data asset leveraging
- Unconstrained by an IT project mindset
- Reporting to the business
- *90 Percent of Large Global Organizations Will Have Appointed Chief Data Officers By 2019*
  (Gartner website accessed January 26, 2016 http://www.gartner.com/newsroom/id/3190117?)

Exorcising the Seven Deadly Data Sins

Context

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges

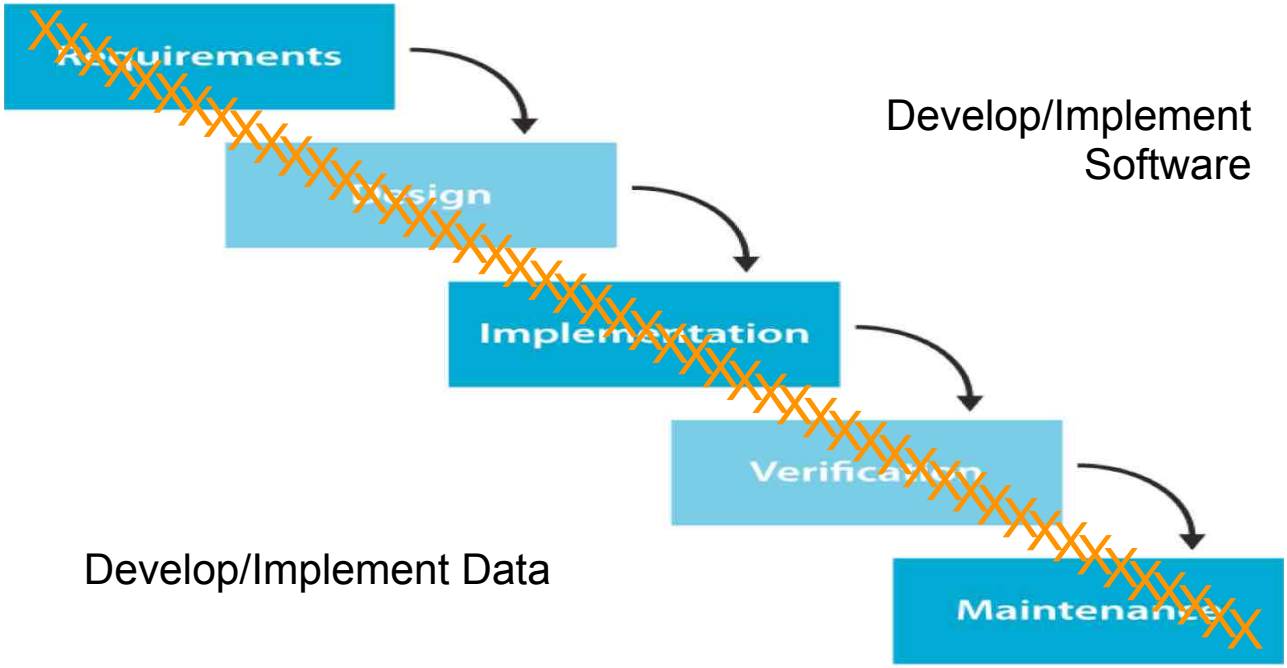


Differences between Programs and Projects

- Programs are Ongoing, Projects End
  - Managing a program involves long term strategic planning and continuous process improvement is not required of a project
- Programs are Tied to the Financial Calendar
  - Program managers are often responsible for delivering results tied to the organization's financial calendar
- Program Management is Governance Intensive
  - Programs are governed by a senior board that provides direction, oversight, and control while projects tend to be less governance-intensive
- Programs Have Greater Scope of Financial Management
  - Projects typically have a straight-forward budget and project financial management is focused on spending to budget while program planning, management and control is significantly more complex
- Program Change Management is an Executive Leadership Capability
  - Projects employ a formal change management process while at the program level, change management requires executive leadership skills and program change is driven more by an organization's strategy and is subject to market conditions and changing business goals

Adapted from http://top.idownloadnew.com/program_vs_project/ and http://management.simplicable.com/management/new/program-management-vs-project-management

Project Implementation

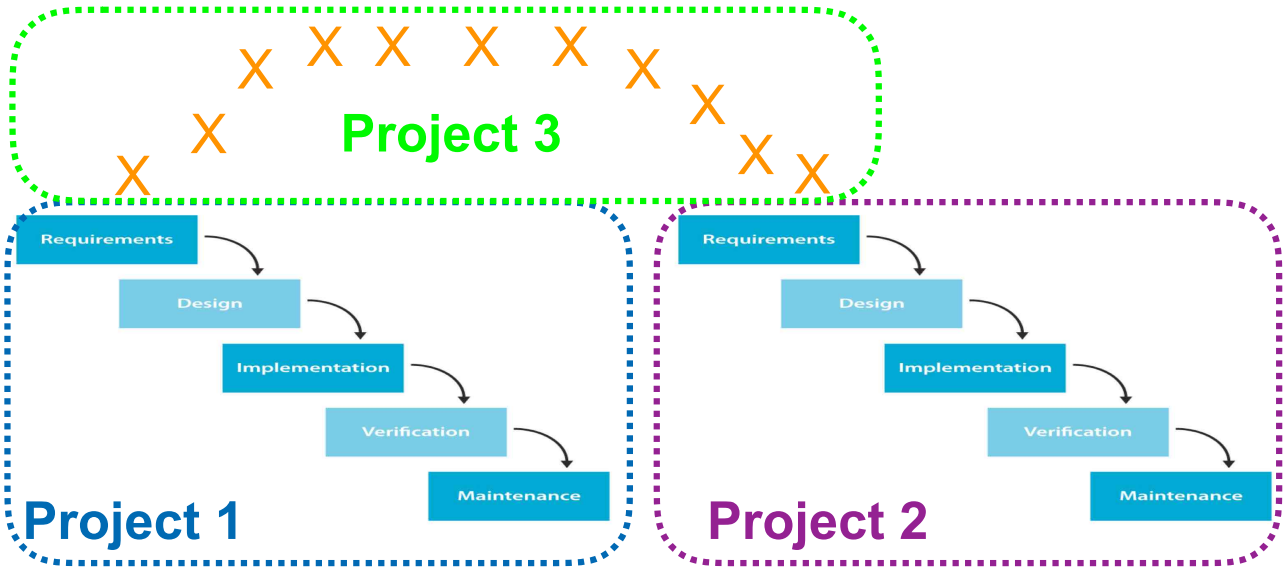


Develop/Implement Software

Develop/Implement Data

This approach can only work when no sharing of data occurs!

Projects Are Silos

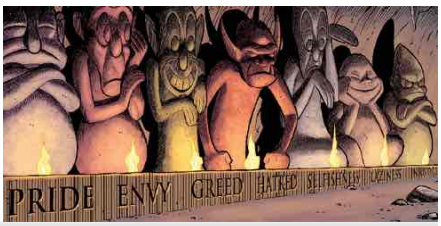


Shared data structures require programmatic development and evaluation

Exorcising the Seven Deadly Data Sins

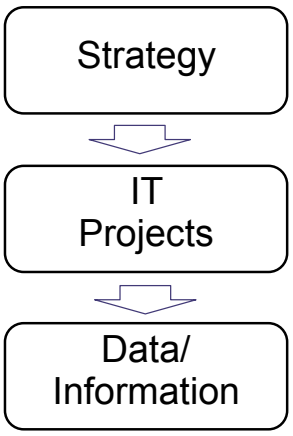
Context

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges



IT Project or Application-Centric Development

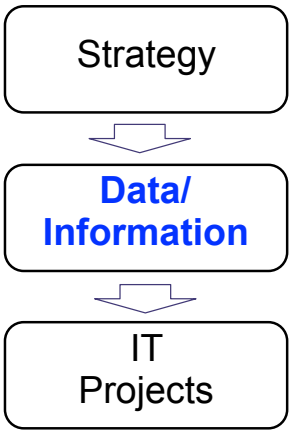
- In support of strategy, organizations implement IT projects
- Data/information are typically considered within the scope of IT projects
- Problems with this approach:
  - Ensures data is formed to the applications and not around the organizational-wide information requirements
  - Process are narrowly formed around applications
  - Very little data reuse is possible



Original articulation from Doug Bagley @ Walmart

Data-Centric Development

- In support of strategy, the organization develops specific, shared data-based goals/objectives
- These organizational data goals/ objectives drive the development of specific IT projects with an eye to organization-wide usage
- Advantages of this approach:
  - Data/information assets are developed from an organization-wide perspective
  - Systems support organizational data needs and compliment organizational process flows
  - Maximum data/information reuse

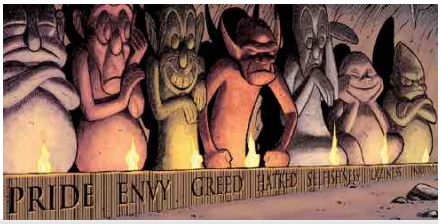


Original articulation from Doug Bagley @ Walmart

Exorcising the Seven Deadly Data Sins

Context

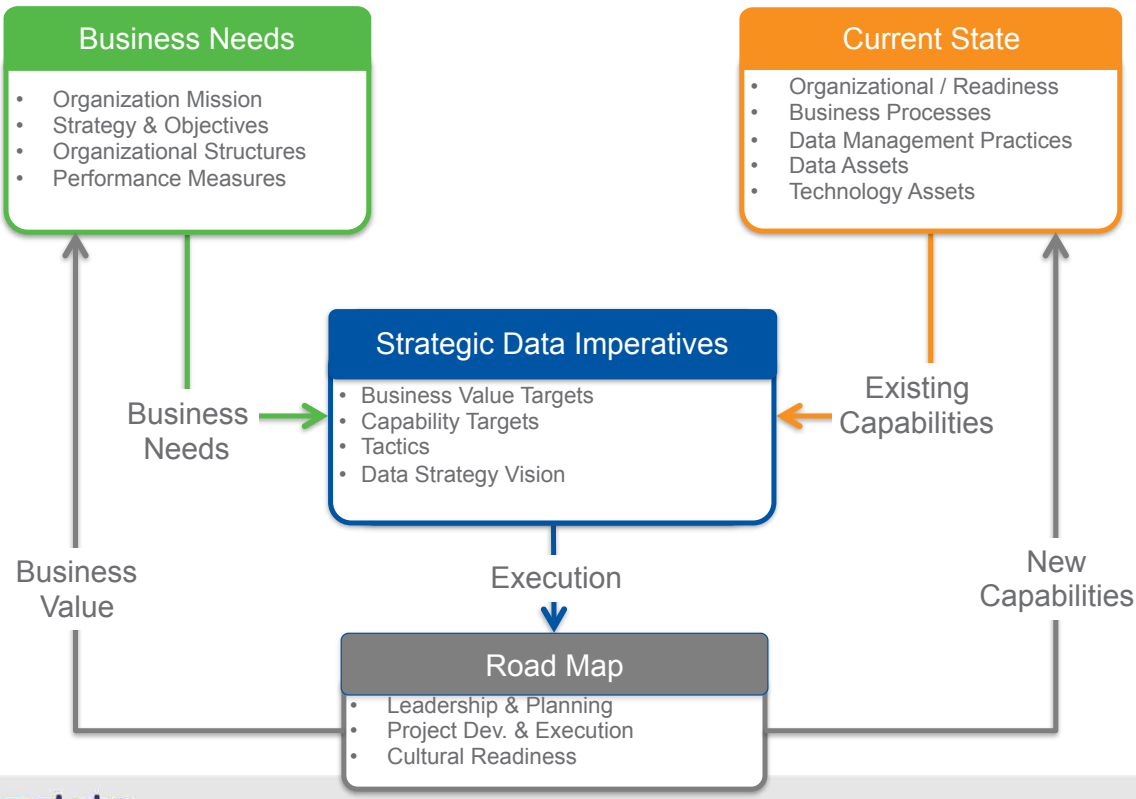
1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges



Data Management Program Expenses

- 5 Data Personel
- $100,000 each annually
- When will you be done?
- "It's okay my CIO gave me 5 years!"

Data Implementation Framework



Virginia Governor's Data Interns Program

gt government technology

MAGAZINE EVENTS PAPERS TOPICS GOVTECH 100

DATA

Virginia Internship Program Pits Grad Students Against Gov Data

For a second year, fresh sets of eyes and cutting-edge data analytics skills are the tools grad students will bring Virginia through the state's data internship program.

BY COLIN WOOD / AUGUST 25, 2015

Virginia Commonwealth University FLICKR/ANDREW BAIN

Commonwealth of Virginia

Office of Governor Terry McAuliffe

For Immediate Release
July 23, 2015

Office of the Governor
Contact: Brian Coy
Email: Brian.Coy@governor.virginia.gov

Governor McAuliffe Announces 2015-16 Data Internships

~ Virginia Commonwealth University graduate student teams to explore the use of data to improve government efficiency ~

RICHMOND – Governor Terry McAuliffe today announced that Virginia state government and the Virginia Commonwealth University School of Business will again work together on data re-engineering internships to explore the use of data to improve the effectiveness and efficiency of state government.

In the 2014-2015 school year, the data internship program's first, 45 graduate students and more than 20 state agencies participated. Those internships have resulted in tangible dollar savings and improved agency processes. Student/agency teams have worked on successful projects, such as

Virginia Secretary of Technology Karen Jackson and CIO of the Commonwealth Nelson Moe are leading the effort on behalf of the state. Students who want to apply for internships should contact Peter Aiken (peter.aiken@vcu.edu) for additional information.

Virginia Natural Heritage Data Explorer

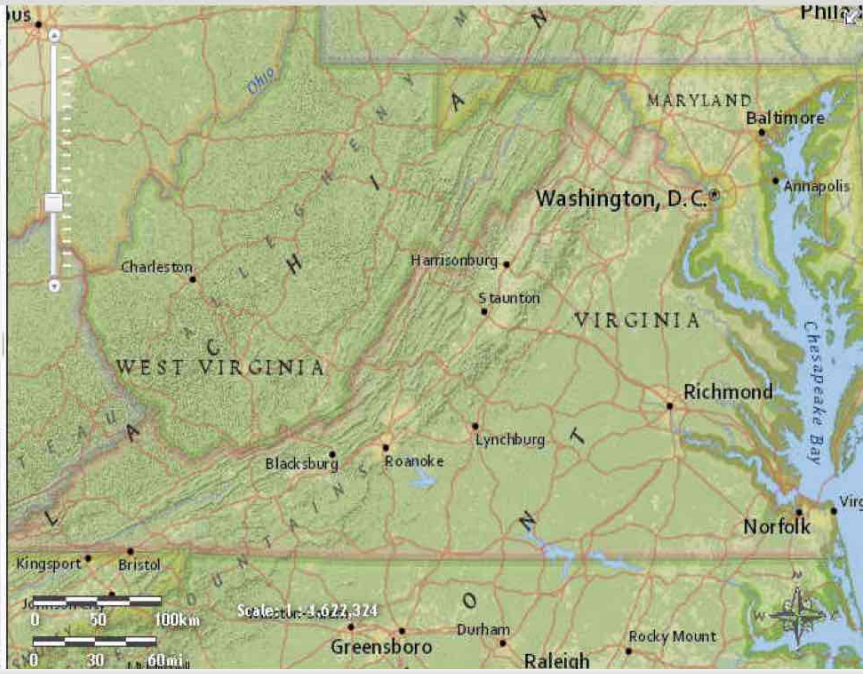
DCR Virginia Department of Conservation & Recreation

Home | Map | About Us | Contact Us | Help | Species/Community Search | Terms and Conditions

Switch Basemap | Add Resource | Find address or place

Layers | Map Making | Find/Results

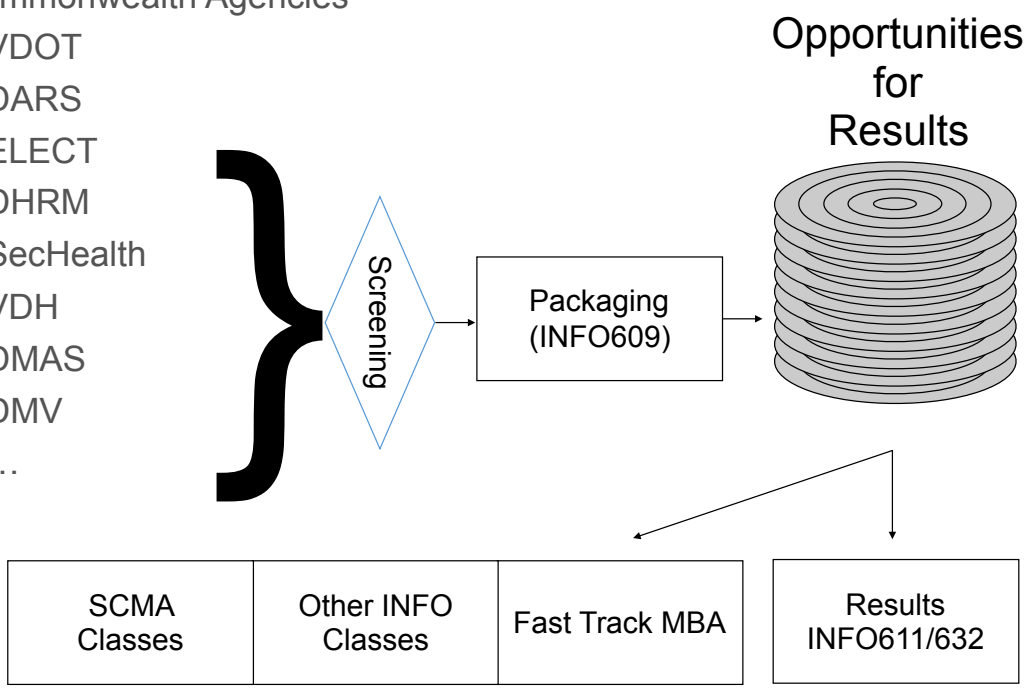
- Managed Conservation Lands
  - Managed Conservation Lands
- Conservation Planning
  - Agricultural Model
  - Cultural Asset Model
  - Ecological Cores
  - Forest Economics Model
  - Recreational Assets Model
  - Vulnerability Model
  - Watershed Integrity Model
- Reference Layers
  - 24K Grid
  - Scenic Rivers
  - Streams (NHD)
  - Trails
  - USGS Placenames
  - VDOT Roads



Two Phase Approach

Commonwealth Agencies

- VDOT
- DARS
- ELECT
- DHRM
- SecHealth
- VDH
- DMAS
- DMV
- …



Hashing Process Illustrated

	First Name	Last Name	Date of Birth	Zip Code	Crime	Sentence	Gender	Race	Disability	LEP	…
1	Peter	Aiken	1/17/1959	23192	A	30	M	W	0	N	
2											
3											

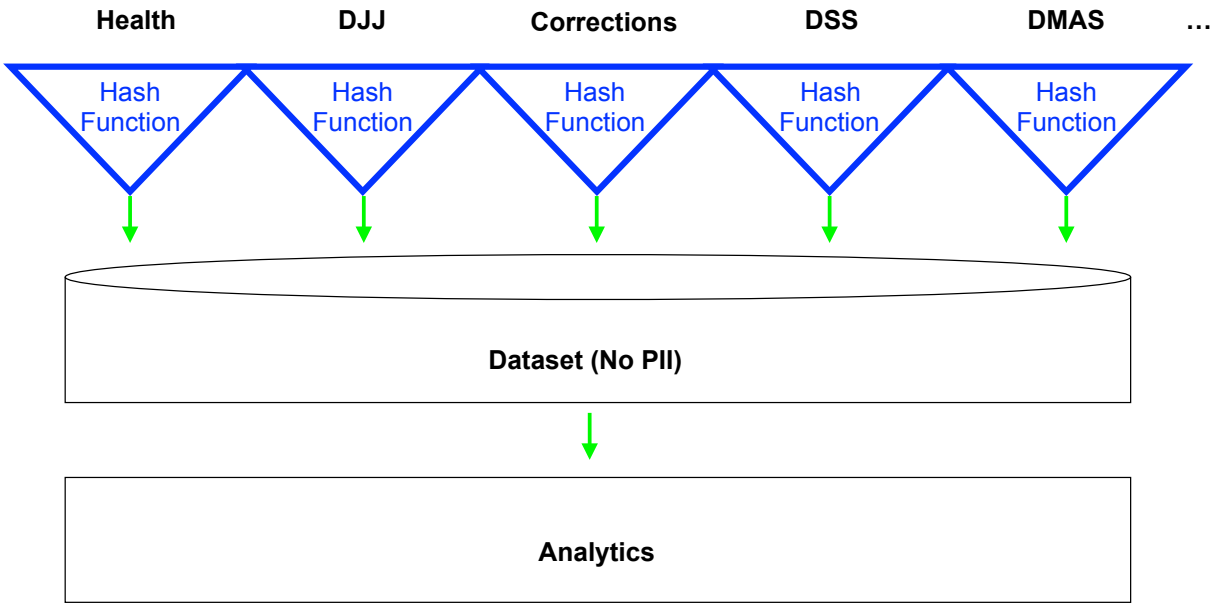
A 30 M W 0 N

Hash Function

(Replaces PII with Anonymous Identifier)

	Anonymous Identifier	Crime	Sentence	Gender	Race	Disability	LEP	…
1	92G-gqP-6ek-oy3	A	30	M	W	0	N	
2								
3								

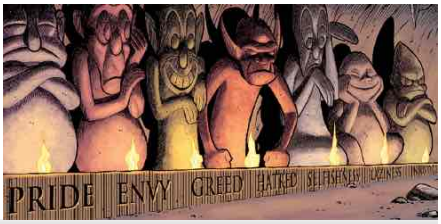
Data Amalgamation Process



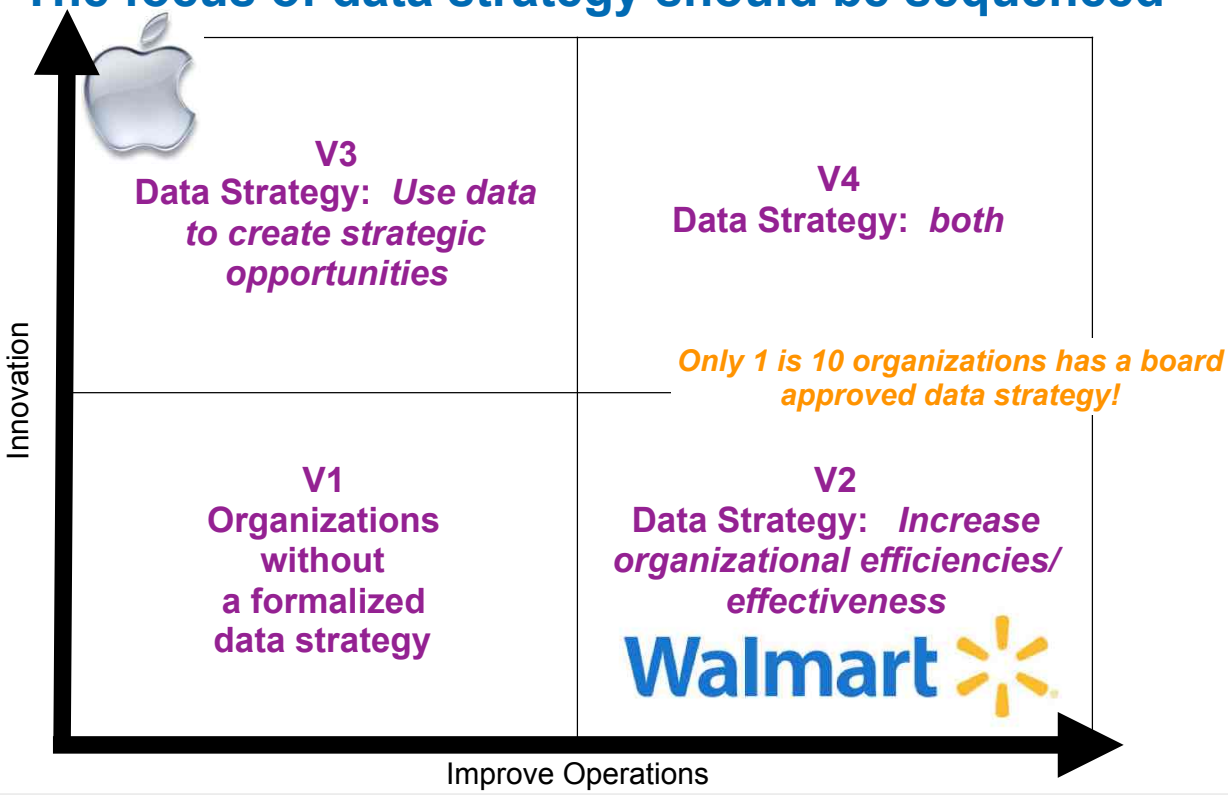
Exorcising the Seven Deadly Data Sins

Context

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges



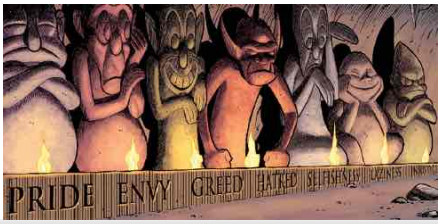
The focus of data strategy should be sequenced



Exorcising the Seven Deadly Data Sins

Context

1. Not Understanding Data-Centric Thinking
2. Lacking Qualified Data Leadership
3. Not implementing a Robust, Programmatic Means of Developing Shared Data
4. Not Aligning The Data Program with IT Projects
5. Failing to Adequately Manage Expectations
6. Not Sequencing Data Strategy Implementation
7. Failing To Address Cultural And Change Management Challenges

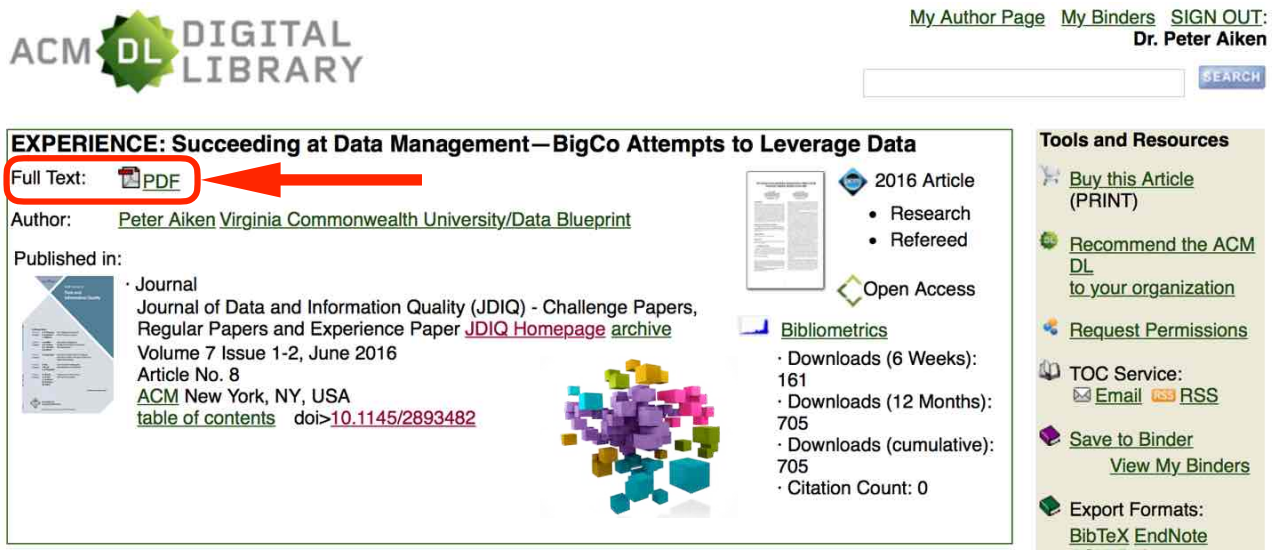


adapted from the *Managing Complex Change* model by Dr. Mary Lippitt, 1987

Diagnosing Organizational Readiness

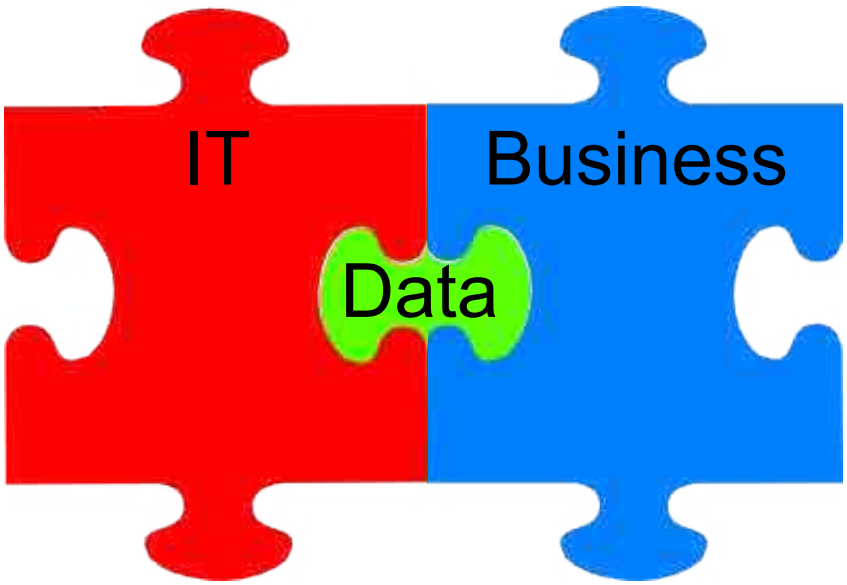
	Skills	Incentive	Resources	Action Plan	Confusion
Vision		Incentive	Resources	Action Plan	Anxiety
Vision	Skills		Resources	Action Plan	Gradual Change
Vision	Skills	Incentive		Action Plan	Frustration
Vision	Skills	Incentive	Resources		False Starts
Vision	Skills	Incentive	Resources	Action Plan	Change

Culture is the biggest impediment to a shift in organizational thinking about data!

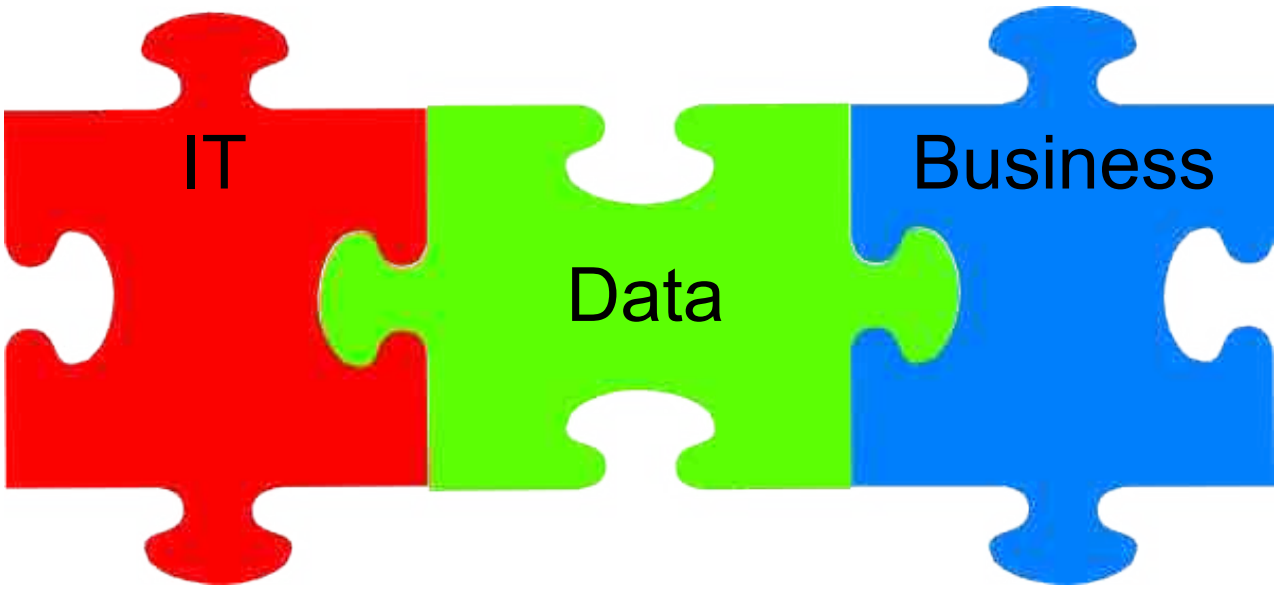


- Free Case Study Download
  - http://dl.acm.org/citation.cfm?doid=2888577.2893482
    or
    http://tinyurl.com/PeterStudy
    or scan the QR Code at the right

As Is State of Data (as Perceived)



Desired To Be State of Data (as Understood)



Wally Easton Playing Piano

UNLOCKING BUSINESS VALUE

Austin

datablueprint.com

10124 W. Broad Street, Suite C
Glen Allen, Virginia 23060
804.521.4056